

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:16:29 ; Search time 91.75 Seconds
(without alignments)
24.287 Million cell updates/sec

Title: US-09-331-631a-1_COPY_117_185
Perfect score: 384
Sequence: 1 NR0RDP00QYEQCQKHCORR.....EEQREDEKYEERMKEDD 69

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 segs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	152.5	39.7	588	VCIB_GOSHI	P09801 gossypium h
2	130.5	34.0	605	VCIB_GOSHI	P09799 gossypium h
3	103.5	27.0	1898	TRHY_HUMAN	Q07283 homo sapien
4	102	26.6	1407	TRHY_HUMAN	P37709 oryctolagus
5	98	25.5	648	KAPC_DICDI	P34099 dictyosteli
6	97.5	25.4	877	INCE_CHICK	P53352 gallus gall
7	96	25.0	2124	Y192_HUMAN	Q93074 homo sapien
8	95	24.7	482	U2R2_HUMAN	Q15696 homo sapien
9	95	24.7	524	SBP_SOYBN	Q04672 glycine max
10	95	24.7	1023	GLT_DROME	P33438 drosophila
11	95	24.5	285	YAFB_SCHPO	Q09663 schizosacch
12	94	24.5	678	INVO_CANFA	P18174 canis fam11
13	94	24.1	538	GARP_PLAUF	P13816 plasmodium
14	92.5	24.0	338	CK11_YEAST	P23291 saccharomyc
15	92	24.0	429	LEGB_PEA	P14594 pisum sativ
16	92	24.0	905	APPA_MACFA	P33621 macaca fasc
17	92	24.0	1403	SNFS_YEAST	P18480 saccharomyc
18	92	24.0	1403	PRO_DROME	P29617 drosophila
19	91.5	23.8	550	BLSA_HUMAN	Q02832 homo sapien
20	91.5	23.8	695	XET_HUMAN	Q02040 homo sapien
21	91	23.7	1344	IF3A_MOUSE	P23116 mus musculu
22	90.5	23.6	758	YM38_YEAST	Q03825 saccharomyc
23	90	23.4	1382	IF3A_HUMAN	Q14152 homo sapien
24	90	23.4	1549	TRHY_SHEEP	P22793 ovis aries
25	90	23.4	1905	TAGB_DICDI	P54683 dictyosteli
26	89.5	23.3	1048	SR4_RAT	Q63627 rattus norv
27	89	23.2	347	INVO_PIG	P18175 sus scrofa
28	89	23.2	743	ABRA_PLAFC	P22620 plasmodium
29	88.5	23.0	573	GLB1_MAIZE	P15590 zea mays (m
30	88.5	23.0	708	GBF_DICDI	P36417 dictyosteli
31	88	22.9	1157	Y182_HUMAN	Q14687 homo sapien
32	88	22.9	1240	YNI1_YEAST	P33935 saccharomyc
33	87.5	22.8	1023	CLOC_DROME	O61735 drosophila

34	86.5	22.5	444	1	CEB_DROME	Q02637 drosophila
35	85	22.1	1178	1	MNN4_YEAST	P36044 saccharomyc
36	84.5	22.0	277	1	TRT1_HUMAN	P13805 homo sapien
37	84.5	22.0	287	1	TRT2_HUMAN	P45179 homo sapien
38	84.5	22.0	389	1	M49_STREX	P15947 streptococ
39	84	21.9	321	1	ABRA_PLAUF	P23746 plasmodium
40	83.5	21.7	1336	1	SEC3_YEAST	P33332 saccharomyc
41	83	21.6	398	1	PF21_ARATH	Q04088 arabidopsis
42	83	21.6	471	1	RU17_XENLA	P09406 xenopus lae
43	83	21.6	544	1	INVO_AOTTR	P24708 lotus trivi
44	83	21.6	644	1	BTD_DROME	Q24266 drosophila
45	82.5	21.5	572	1	MOES_LYTVA	P52962 lytechinus

ALIGNMENTS

```
RESULT 1
ID VCIB_GOSHI STANDARD; PRT; 588 AA.
AC P09801;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE VICILIN C72 PRECURSOR (ALPHA-GLOBULIN B).
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Malvales; Malvaceae; Gossypium.
RN (1)
RP SEQUENCE FROM N.A.
RA Chlan C.A., Pyle J.B., Legocki A.B., Dure L. III;
RT "Developmental biochemistry of cottonseed embryogenesis and
RT germination. XVII. cDNA and amino acid sequences of the members of
RT the storage protein families.";
RL Plant Mol. Biol. 7:475-489(1986).
CC -!- FUNCTION: SEED STORAGE PROTEIN.
CC -!- SUBCELLULAR LOCATION: CYTOSOL; MEMBRANE-BOUND VACUOLAR PROTEIN
CC BODIES.
CC -!- SIMILARITY: NO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC CONVICTILIN, CONGLICTININ, ETC.).
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: M16891; AAA33071.1; -
CC PIR: A30838; FMCNAB.
CC DR HSSP: P50477; ICAN.
CC DR INTERPRO: IPR001113; -
CC DR PFAM: PF00546; Seedstore_7s; 1.
CC Seed storage protein; Signal.
CC FT SIGNAL 1 25
CC CHAIN 26 588 VICILIN C72
CC SEQUENCE 588 AA; 69729 MW; 63E699E29AB8ADEB CRC64;
SO
Query Match 39.7%; Score 152.5; DB 1; Length 588;
Best local Similarity 34.8%; Pred. No. 7.3e-06;
Matches 31; Conservative 20; Mismatches 15; Indels 23; Gaps 3;
```

```

RESULT 2
VCLA_GOSHI STANDARD: PRT: 605 AA.
ID VCLA_GOSHI
AC P09799:
DT 01-MAR-1989 (rel. 10, Created)
DT 01-MAR-1989 (rel. 10, Last sequence update)
DT 15-JUL-1999 (rel. 38, Last annotation update)
DE VICILIN GC72-A PRECURSOR (ALPHA-GLOBULIN A) .
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons, core eudicots; Rosidae; eurosids II;
OC Malvales; Malvaceae; Gossypium.
RN [1]
RP SEQUENCE FROM N.A.
RA Chlan C.A., Borroto K., Kamalay J.A., Dure L. III;
RT "Developmental biochemistry of cottonseed embryogenesis and
germination. XIX. Sequences and genomic organization of the alpha
globulin (vicillin) genes of cottonseed."
RL Plant Mol. Biol. 9:533-546(1987).
CC -1- FUNCTION: SEED STORAGE PROTEIN.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC MEMBRANE-BOUND VACUOLAR PROTEIN
CC BODIES.
CC -1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC CONVICTILIN, CONGLYCININ, ETC.).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M19378; AAA3069.1; -.
DR PIR: S06398; S06398.
DR HSP: P50477; ICAX.
DR INTERPRO: IPR001113; -.
DR PFM: PF00546; Seedstore_7s; 1.
DR KW Seed storage protein; Signal.
FT SIGNAL 1 23
FT CHAIN 24 605 VICILIN GC72-A.
SQ SEQUENCE 605 AA: 71049 MW: C9DB9371C976553B CRC64:

```

```

RA Lee S.-C., Kim I.-G., Marekov L.N., O'Keefe E.J., Parry D.A.D.,
RA Steiner P.M.;
RT "The structure of human trichohyalin. Potential multiple roles as a
RT functional EF-hand-like calcium-binding protein, a confined cell
RT envelope precursor, and an intermediate filament-associated (cross-
RT linking) protein."
RL J. Biol. Chem. 268:12164-12176(1993).
RN [2]
RP SEQUENCE OF 1731-1898 FROM N.A., AND CHARACTERIZATION.
RX MEDLINE-93315897; PubMed-7686953;
RA O'Keefe E.J., Hamilton E.H., Lee S.-C., Steiner P.M.;
RT "Trichohyalin: a structural protein of hair, tongue, nail, and
RT epidermis."
RL J. Invest. Dermatol. 101:655-715(1993).
CC -1- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
CC IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
CC INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
CC LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
CC ISODIPEPTIDE BONDS. IT MAY SERVE AS SCARFOLD PROTEIN, TOGETHER
CC WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN
CC ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN
CC ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
CC DIFFERENTIATION.
CC -1- SUBUNIT: MONOMER (PROBABLE).
CC -1- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS
CC THE INNER ROOT SHEATH (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN
CC THE FILIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF
CC THE EPIDERMIS.
CC -1- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND
CC CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST
CC ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS
CC OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRANDED
CC ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS
CC THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.
CC DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN
CC THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG
CC DIFFERENT SPECIES.
CC -1- PTM: KNOWN SUBSTRATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE
CC PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMINASE.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE S-100
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: L09190; AAA65582.1; -.
DR PIR: A45973; A45973.
DR HSP: F02633; 1B0C.
DR MIM: 190370; -.
DR INTERPRO: IPR001751; -.
DR INTERPRO: IPR002048; -.
DR PFM: PF01023; S_100; 1.
DR PFM: PF00036; eHand; 1.
DR PROSITE: PS00018; EF_HAND; 1.
DR PROSITE: PS00303; S100_CABP; 1.
DR Repeat; Calcium-binding.
FT DOMAIN 1 91
FT CA_BIND 22 33 SITE I (LOW AFFINITY) (POTENTIAL).
FT CA_BIND 62 73 SITE II (HIGH AFFINITY) (POTENTIAL).
FT DOMAIN 314 390 6 X 13 AA TANDEM REPEATS OF
FT REPEAT 314 326 R-R-E-Q-E-E-E-R-R-E-Q-Q-L.
FT REPEAT 327 339 1-1 (APPROXIMATE).
FT REPEAT 340 351 1-2 (APPROXIMATE).
FT REPEAT 352 364 1-3 (APPROXIMATE).
FT REPEAT 365 377 1-4.
FT REPEAT 1-5.

```

```

FT REPEAT 378 390 1-6.
FT DOMAIN 391 444 9 X 6 AA TANDEM REPEATS OF R-R-E-Q-Q-L.
FT REPEAT 391 444 2-1.
FT REPEAT 397 402 2-2.
FT REPEAT 403 408 2-3.
FT REPEAT 409 414 2-4.
FT REPEAT 415 420 2-5.
FT REPEAT 421 426 2-6.
FT REPEAT 427 432 2-7.
FT REPEAT 433 438 2-8.
FT REPEAT 439 444 2-9.
FT DOMAIN 444 702 9 X 28 AA APPROXIMATE TANDEM REPEATS.
FT DOMAIN 923 952 8 X 30 AA TANDEM REPEATS.
FT REPEAT 923 952 4-1.
FT REPEAT 953 982 4-2.
FT REPEAT 1013 1042 4-3.
FT REPEAT 1043 1072 4-4.
FT REPEAT 1073 1102 4-5.
FT REPEAT 1103 1132 4-6.
FT REPEAT 1133 1162 4-7.
FT DOMAIN 1250 1849 23 X 26 AA APPROXIMATE TANDEM REPEATS.
FT CONFLICT 1752 1752 F -> L (IN REF. 2).
FT CONFLICT 1794 1801 Q -> K (IN REF. 2).
FT CONFLICT 1857 1857 V -> G (IN REF. 2).
FT CONFLICT 1880 1880
SQ SEQUENCE 1898 AA; 247219 MW; A74B5947FB62E31D CRC64;

```

Query Match 27.0%; Score 103.5; DB 1; Length 1898;
 Best Local Similarity 36.2%; Pred. No. 0.17;
 Matches 25; Conservative 19; Mismatches 18; Indels 7; Gaps 3;

```

OY 2 RQDPQOQYE-QCKHCRRETEPRHMQTCQRCERYEKEKRRKQKRYEEO--REDDEE 58
DB 290 RQEEQOQQRKRRKQOQLRRKQOEBRR---EQOEBRRQOQRRQOEBRRQOQLRRQOE 345
OY 59 KYERKKEE 67
DB 346 RREQQLRRE 354

```

```

RESULT 4
TRAY_RABIT STANDARD; PRT; 1407 AA.
AC P37709;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE TRICHOHYALIN.
GN THH.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN (1)
RP SEQUENCE FROM N.A.
RA Fietz M.J., Rogers G.E.;
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
CC IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
CC INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
CC LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
CC ISOPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER
CC WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN
CC ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN
CC ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
CC DIFFERENTIATION.
CC -1- SUBUNIT: HOMODIMER (PROBABLE).
CC -1- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS
CC THE INNER ROOT SHEAT (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN
CC THE FILIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF
CC THE EPIDERMIS.

```

```

CC -1- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND
CC CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST
CC ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS
CC OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRANDED
CC ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS
CC THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.
CC DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN
CC THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG
CC DIFFERENT SPECIES.
CC -1- PTM: KNOWN SUBSTRATE OF TRANSGUTAMINASE. SOME 200 ARGININES ARE
CC PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMINASE.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO THE S-100
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z19092; CAA79519.1; -.
CC DR PIR; S28589; S28589.
CC DR HSSP; P02633; 1BOC.
CC DR INTERPRO; IPR001751; -.
CC DR INTERPRO; IPR002048; -.
CC DR PRAM; PRF01023; S_100; 1.
CC DR PRAM; PRF00036; efhand; 1.
CC DR PROSITE; PS00018; EF_HAND; 1.
CC DR PROSITE; PS00303; S100_CABP; 1.
CC KW Repeat: Calcium-binding.
CC FT DOMAIN 1 91 S-100 LIKE.
CC FT CA_BIND 22 33 SITE 1 (LOW AFFINITY) (POTENTIAL).
CC FT CA_BIND 62 73 SITE 11 (HIGH AFFINITY) (POTENTIAL).
CC SQ SEQUENCE 1407 AA; 183781 MW; AE17D2A159F12B7F CRC64;

```

Query Match 26.6%; Score 102; DB 1; Length 1407;
 Best Local Similarity 33.8%; Pred. No. 0.17;
 Matches 25; Conservative 20; Mismatches 19; Indels 10; Gaps 2;

```

OY 3 QRPDQOQYE-QCKHCRRETEPRHMQTCQRCERYEKEKRRKQKRYEEO--REDDEE 54
DB 256 QOQLRRLEERIRERQOERREDO--QLRRQOERLEERERQOQLRRLEERIRERQRL 313
OY 55 EDEKYEERKKEED 68
DB 314 EOEBRRRQRLQOEE 327

```

```

RESULT 5
KAPC_DICDI STANDARD; PRT; 648 AA.
ID KAPC_DICDI
AC P34099;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CAMP-DEPENDENT PROTEIN KINASE CATALYTIC SUBUNIT* (EC 2.7.1.37).
GN PKAC OR PK2 OR PK3.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Dictyostellida; Dictyostellium.
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=91323730; PubMed=1864510;
RA Buerki E., Anjard C., Scholder J.C., Reymond C.D.;
RT *Isolation of two genes encoding putative protein kinases regulated
RT during Dictyostelium discoideum development.*;
RL Gene 102:57-65(1991).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=93385090; PubMed=8373760;

```



```

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BONE MARROW;
RX MEDLINE=96281124; PubMed=8724849;
RA Nagase T., Seki N., Ishikawa K.-I., Tanaka A., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. V.
RT The coding sequences of 40 new genes (KIA0161-KIA0200) deduced by
RT analysis of cDNA clones from human cell line KG-1."
RL DNA Res. 3:17-24(1996).
CC -1- TISSUE SPECIFICITY: UBIQUITOUS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D83783; BAA12112.1; -.
KM Hypothetical protein.
FT NON-TER 1
FT DOMAIN 599 602 POLY-SER.
FT DOMAIN 1201 1207 POLY-GLY.
FT DOMAIN 1998 2124 GLN-RICH.
FT DOMAIN 1998 2023 POLY-GLN.
FT DOMAIN 2028 2033 POLY-GLN.
FT DOMAIN 2037 2070 POLY-GLN.
FT DOMAIN 2090 2097 POLY-GLN.
SQ SEQUENCE 2124 AA; 237207 MW; 255FB9419EC39FA2 CRC64;

Query Match 25.0%; Score 96; DB 1; Length 2124;
Best Local Similarity 21.4%; Pred. No. 0.75;
Matches 15; Conservative 35; Mismatches 16; Indels 4; Gaps 1;

QY 2 RORPQOQYECCKHCKRREPRHMOTCOO---RCRRYRKRRKROKRYEQRDEDE 57
   11 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 2001 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ 2060

QY 58 EKYERKKEE 67
   : : : : :
Db 2061 QQHQQQQQQQQ 2070

RESULT 8
ID U2R2_HUMAN STANDARD; PRT; 482 AA.
AC 015696;
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE U2 SMALL NUCLEAR RIBONUCLEOPROTEIN AUXILIARY FACTOR 35 KDA SUBUNIT
DE RELATED-PROTEIN 2.
GN U2AF1-RS2 OR U2AF1RS2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRNIN;
RX MEDLINE=96163878; PubMed=8586425;
RA Kitagawa K., Wang X., Hatada T., Yamaoka T., Nojima H.,
RA Inazawa J., Abe T., Mitsuya K., Oshimura M., Murata A., Monden M.,
RA Mukai T.;
RT "Isolation and mapping of human homologues of an imprinted mouse gene
RT U2af1-rs1."
RL Genomics 30:257-263(1995).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).

```

```

CC -1- SIMILARITY: TO MAMMALIAN SPLICING FACTOR U2AF 35 KDA SUBUNIT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D49677; BAA08533.1; -.
DR MIM: 300028; -.
DR INTERPRO: IPR000504; -.
DR INTERPRO: IPR000571; -.
DR PFAM: PF00076; rrm; 1.
DR PFAM: PF00642; zf-CCCH; 1.
DR PROSITE: PS0102; RRM; 1.
KW Nuclear protein; Ribonucleoprotein.
FT DOMAIN 46 49 POLY-GLU.
FT DOMAIN 118 123 POLY-GLU.
SQ SEQUENCE 482 AA; 58044 MW; 1DACCB8A6CA4727A6 CRC64;

Query Match 24.7%; Score 95; DB 1; Length 482;
Best Local Similarity 27.7%; Pred. No. 0.25;
Matches 18; Conservative 26; Mismatches 21; Indels 0; Gaps 0;

QY 3 QRPQOQYECCKHCKRREPRHMOTCOOCCRRYRKRRKROKRYEQRDEDEKYE 62
   11 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 65 ERERQRLHEWLLREQAKQAEFRKKEKBAARKROEDERLUKEWEDQRREREDEQ 124

QY 63 RKKEE 67
   : : : : :
Db 125 KRQEK 129

RESULT 9
ID SBP_SOYBN STANDARD; PRT; 524 AA.
AC 004672;
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE SUCROSE-BINDING PROTEIN PRECURSOR (SBP).
DE SBP.
GN Glycine max (Soybean).
OS Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fabales; Fabaceae; Papilionoideae; Glycine.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 30-50.
RC TISSUE=EMBRYO;
RX MEDLINE=93104680; PubMed=1467654;
RA Grimes H.D., Overvoorde P.J., Ripp K., Franceschi V.R., Hiltz W.D.;
RT "A 62-kD sucrose binding protein is expressed and localized in
RT tissues actively engaged in sucrose transport."
RL Plant Cell 4:1561-1574(1992).
CC -1- FUNCTION: PLAYS A ROLE IN SUCROSE TRANSPORT.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED.
CC -1- TISSUE SPECIFICITY: ASSOCIATED WITH THE PLASMA MEMBRANE OF
CC SEVERAL CELL TYPES ENGAGED IN SUCROSE TRANSPORT, INCLUDING THE
CC MESOPHYLL CELLS OF YOUNG SINK LEAVES, THE COMPANION CELLS OF
CC MATURE PHLOEM AND THE CELLS OF DEVELOPING COTYLEDONS.
CC -1- DEVELOPMENTAL STAGE: IN THE COTYLEDON, EXPRESSION IS NOT DETECTED
CC UNTIL 10 DAYS AFTER FERTILIZATION. BETWEEN 10-19 DAYS AFTER
CC FERTILIZATION, EXPRESSION INCREASES RAPIDLY BUT DECLINES 20-30
CC DAYS AFTER FERTILIZATION. 30 DAYS AFTER FERTILIZATION, NO
CC EXPRESSION OCCURS. THIS EXPRESSION PATTERN CLOSELY PARALLELS THE
CC RATE OF SUCROSE UPTAKE IN THE COTYLEDON.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L06038; AAB03894.1; -.
DR PIR: J01730; J01730.
DR HSSP: P50477; 1CAM.
DR INTERPRO: IPR001113; -.
DR PFAM: PF00546; Seedstore.7s; 1.
KW Transport; Sugar transport; Signal; Membrane.
FT SIGNAL 1 29
FT CHAIN 30 524
FT SEQUENCE 524 AA; 60522 MW; 0251EE90796EF341 CRC64;

Query Match
Best Local Similarity 35.9%; Score 95; DB 1; Length 524;
Matches 23; Conservative 15; Mismatches 16; Indels 10; Gaps 3;

QY 12 QCGHCORRETEPRHNOTCOQRCER--RYEKERKQ-----QKRYEQORDEDEKYEERM 64
DB 46 QCGQDQDQYTGDKR---VCLQSDRTHRMKQREKQIQEETREKKEEESRERERFQDQEH 102
QY 65 KEED 68
DB 103 EEDD 106

RESULT 10
GLT_DROME STANDARD; PRT; 1023 AA.
ID GLT_DROME
AC P33438;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE GLUTACTIN PRECURSOR.
GN GLT.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=OREGON-R;
RX MEDLINE=90214632; PubMed=2108664;
RA Olson P.F., Fessler L.I., Nelson R.E., Sterne R.E., Campbell A.G.,
RA Fessler J.H.;
RT "Glutactin, a novel Drosophila basement membrane-related glycoprotein
RT with sequence similarity to serine esterases.";
RL EMBL J. 9:1219-1227(1990).
CC -1- FUNCTION: NOT KNOWN. BINDS CALCIUM IONS.
CC -1- SUBCELLULAR LOCATION: BASEMENT MEMBRANES.
CC -1- PTM: EXTENSIVELY O-GLYCOSYLATED AND ALSO N-GLYCOSYLATED.
CC -1- PTM: FOUR TYROSINES ARE SULFATED.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO THE TYPE-B
CC CARBOXYLESTERASE/LIPASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X53286; CAA37380.1; -.
DR HSSP: F21836; 1MAH.
DR FLVBASE: FBgn000114; GLT.
DR INTERPRO: IPR002018; -.
DR PFAM: PF00135; Coesterase; 2.
DR PROSITE: PS00941; CARBOXYLESTERASE_B_2; 1.
KW Glycoprotein; Sulfatation; Calcium-binding; Signal.

```

```

FT SIGNAL 1 17
FT CHAIN 18 1023
FT SIMILAR 18 602
FT DOMAIN 603 615
FT DOMAIN 616 1023
FT CARBOHYD 115 115
FT CARBOHYD 368 368
FT CARBOHYD 402 402
FT CARBOHYD 810 810
FT DISULFID 123 145
FT DISULFID 298 316
FT SEQUENCE 1023 AA; 118412 MW; 3638CF9AB60E8C CRC64;

Query Match
Best Local Similarity 34.7%; Score 95; DB 1; Length 1023;
Matches 26; Conservative 19; Mismatches 22; Indels 8; Gaps 3;

QY 1 NRGDRPQQ--YBQCGHCORRETEPRHNOTCOO---RCRRYERKRRKQKR--YEED 52
DB 671 DQQRYPQOEPRQDQDRIRQQRQOEERLRQQRQOEERLRQQRQOEERLRQQRQOEYERE 730
QY 53 QRDEDEKYEERKKEE 67
DB 731 QQRQERREERERQ 745

RESULT 11
YAFASCHPO STANDARD; PRT; 1085 AA.
ID YAFASCHPO
AC Q09863;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHETICAL 122.9 KDA PROTEIN C29E6.10C IN CHROMOSOME I.
GN SPAC29E6.10C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
OC Schizosaccharomycetaceae; Schizosaccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Jones L., Murphy L., McNeil A., Simpson I., Harris D., Barrell B.G.,
RA Rajandream M.A., Walsh S.V.;
RL Submitted (OCT-1995) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: TO YEAST YN1091W.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z66525; CAA91432.1; -.
DR Hypothetical protein.
FT DOMAIN 33 46
FT DOMAIN 184 194
FT DOMAIN 695 698
FT DOMAIN 714 717
FT DOMAIN 718 721
FT DOMAIN 938 941
FT SEQUENCE 1085 AA; 122940 MW; 041164132676F233 CRC64;

Query Match
Best Local Similarity 29.1%; Score 95; DB 1; Length 1085;
Matches 23; Conservative 22; Mismatches 20; Indels 14; Gaps 2;

QY 2 RQRDPQQYEQG-QKHQCRRETE-----PRHMQCOQRCRRYERKRRKQK 47
DB 596 KRDKRKQKLAKKEERQREARLAEQAQAKALEAKRKQEAARKKREQRRLKREDEKKQ 655

```

```

OY 48 RYEQQRDEKRYERME 66
Db 656 ELEROKREKOKERREK 674

RESULT 12
INVO_CANFA STANDARD; PRT; 285 AA.
ID INVO_CANFA
AC P18174;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE INVOLUCRIN.
GN IVL.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eumleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
RN
RP
SEQUENCE FROM N.A.
RA MEDLINE=90348475; PubMed=2385171;
RX Tseng H., Green H.;
RT "The involucrin genes of pig and dog: comparison of their segments of
RL Mol. Biol. Evol. 7:293-302(1990).
CC -1- FUNCTION: INVOLUCRIN IS A KERATINOCYTE PROTEIN THAT FIRST APPEARS
CC IN THE CELL CYTOSOL, BUT ULTIMATELY BECOMES CROSS-LINKED TO
CC MEMBRANE PROTEINS BY TRANSGUTAMINASE. ALL THAT RESULTS IN THE
CC FORMATION OF AN INSOLUBLE ENVELOPE BENEATH THE PLASMA MEMBRANE.
CC -1- TISSUE SPECIFICITY: PRESENT IN KERATINOCYTES OF EPIDERMIS AND
CC OTHER STRATIFIED SQUAMOUS EPITHELIA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M34442; AAA30853.1;
DR INTERPRO: IPR002360;
DR PROSITE: PS00795; INVOLUCRIN; 1.
KW Keratinocyte; Repeat.
SQ SEQUENCE 285 AA; 33384 MW; DCE1BD88B9248BEA CRC64;

Query Match 24.5%; Score 94; DB 1; Length 285;
Best Local Similarity 31.1%; Pred. No. 0.19;
Matches 23; Conservative 19; Mismatches 22; Indels 10; Gaps 2;

OY 2 RORPQOQYEQCCKRREPRHMOTCOQRCRERYEKERKROKRYEQQRDEE---- 57
Db 166 QQQPQQGKQKQCEKHQAKNLEQDLQGAQR--KQQKQKLEQEKRLVDQHLQDEPQQR 223
OY 58 ----EYERMEKEE 67
Db 224 TEQEKKEQVLEQ 237

RESULT 13
GARP_PLAUF STANDARD; PRT; 678 AA.
ID GARP_PLAUF
AC P13816;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GLUTAMIC ACID-RICH PROTEIN PRECURSOR.
GN GARP.
OS Plasmodium falciparum (Isolate FC27 / Papua New Guinea).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
RN
RP
SEQUENCE FROM N.A.

```

```

RX MEDLINE=89040048; PubMed=2903445;
RA Triglia T., Stahl H.-D., Crewther P.E., Silva A., Anders R.F.,
RA Kemp D.J.;
RT "Structure of a Plasmodium falciparum gene that encodes a glutamic
RT acid-rich protein (GARP)".
RL Mol. Biochem. Parasitol. 31:199-202(1988).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: J03998; AAA29605.1;
DR PIR: A54514; A54514.
KW Repeat; Malaria; Antigen; Signal.
FT SIGNAL
FT CHAIN 26 678
FT DOMAIN 120 164 15 X 3 AA TANDEN REPEATS OF K-K-X.
FT DOMAIN 372 416 9 X APPROXIMATE TANDEN REPEATS.
FT DOMAIN 417 441 5 X APPROXIMATE TANDEN REPEATS.
FT DOMAIN 576 604 POLY-GLU.
FT DOMAIN 605 653 7 X APPROXIMATE TANDEN REPEATS.
FT DOMAIN 654 663 POLY-GLU.
SQ SEQUENCE 678 AA; 80551 MW; 2A8F85606496EA9E CRC64;

Query Match 24.5%; Score 94; DB 1; Length 678;
Best Local Similarity 24.6%; Pred. No. 0.4;
Matches 17; Conservative 28; Mismatches 20; Indels 4; Gaps 1;

OY 3 QRPDQOQYEQC---QKHCORREPRHMOTCOQRCRERYEKERKROKRYEQQRDEE 58
Db 247 KKHQOQEMKTLTKRERKQKEMKEQEKTEKKRQKQKQKQKQKQKQK 306
OY 59 KYERMEKEE 67
Db 307 QKEREMKKQ 315

RESULT 14
CK11_YEAST STANDARD; PRT; 538 AA.
ID CK11_YEAST
AC P23291;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE CASEIN KINASE I HOMOLOG 1 (EC 2.7.1.1).
GN YCK1 OR CK12 OR YHR135C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN
RP
SEQUENCE FROM N.A.
RC STRAIN=5288C;
RX MEDLINE=92108037; PubMed=1729698;
RA Robinson L.C., Hubbard E.J.A., Graves P.R., de Paoli-Roach A.A.,
RA Roach P.J., Kung C., Haas D.W., Hagedorn C.H., Goebel M.,
RA Culbertson M.R., Carlson M.;
RT "Yeast casein kinase I homologues: an essential gene pair.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:28-32(1992).
RN
RP
SEQUENCE FROM N.A.
RX MEDLINE=92329995; PubMed=1627830;
RT Wang P.-C., Vancura A., Mitcheson T.G.M., Kuret J.;
RT "Two genes in Saccharomyces cerevisiae encode a membrane-bound form
RL of casein kinase-1."
RN Mol. Biol. Cell 3:275-286(1992).
RN
RP
SEQUENCE FROM N.A.
RC STRAIN=5288C / AB972;

```

